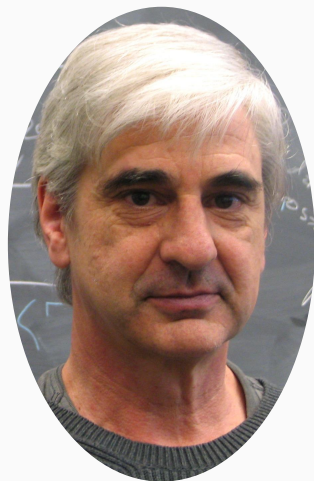


Celebrate Learning Panel

Creating Inclusion and Accessibility through Data Science: Challenges and Solutions

13 May 2022

Presenters



Phil Austin

*Earth, Ocean and
Atmospheric Science*

(UBC-V)



Jonathan Graves

*Vancouver School of
Economics*

(UBC-V)



Melissa Lee

Statistics

(UBC-V)



Stephen Michaud

*Centre for Teaching, Learning
and Technology*

(UBC-V)



Firas Moosvi

*Computer Science,
Mathematics, Physics, and
Statistics*

(UBC-O)

Land Acknowledgement



Motivation

- Data science has exploded in popularity, transforming many disciplines
- It has the potential to radically democratize access to information

- However, the benefits of this revolution have tended to accumulate within the parts of society most well-situated to adopt them

- What new challenges does this new paradigm create for inclusion and access?
- What technologies, tools, techniques, and pedagogies can we use to help mitigate or overcome these challenges?

Outline

- Introductions
- Challenges
 - Tech stack and introducing data science
 - Diverse, challenging IT landscape
 - Barriers to participation and understanding in courses
 - Assessment, especially at scale
 - Hardware and research costs
- Opportunities
 - JupyterBook and JupyterLite
 - Centralized LT support
 - Open source teaching
 - Testthat and formative feedback
 - Cloud-based tools and frameworks
- Q&A

Challenges

Teaching students data science tools with varied technical proficiency

GETTING STARTED!

Data Science Tech Stack

macOS

Windows

☰ Contents

Installation notes

Table of Contents

UBC Student Email

Web browser

Zoom

GitHub.com account

Microsoft Office

Python

[Visual Studio Code](#)

Terminal

Git and GitHub

Tree

Test JupyterLab

Teaching students data science tools with varied technical proficiency

GETTING STARTED!

Data Science Tech Stack ^

macOS

Windows

☰ Contents

Installation notes

Table of Contents

UBC Student Email

Web browser

Zoom

GitHub.com account

Microsoft Office

Python

[Visual Studio Code](#)

Terminal

Git and GitHub

Tree

Test JupyterLab

Data 301: Rate your level of programming experience coming into the course

185 Responses



Student and Instructor Learning Technology Support

- Learning Technology Hub (LT Hub) receives difficult support requests:
 - Installation of software on laptops, tablets, and other devices
 - Availability of software packages and tools on systems not centrally supported
 - Many and varied compute requests
- Pilot project to develop and determine the best approach to supporting computational needs in an accessible and equitable fashion



EOSC 211: intro programming course

- **Economic/logistical barriers**
 - Taught in Matlab – free for students, but minimum \$900 US/year for a basic license after graduation (plus \$440 US/year per toolbox)
 - implicit requirement for a laptop to work outside of building hours
 - Limited number of open seats for drop-ins for the EOAS computer lab

- **Instructional materials**
 - Textbook (\$80 for paperback version), free e-book access with UBC cwl
 - All instructional material (powerpoint, code) in canvas silo
 - No way for transfer students or others to inspect/prepare
 - No access to the course materials after course completion

DSCI100: Challenges with assessment

- Scaling grading for large courses
- Providing high quality consistent qualitative feedback can be challenging
- Haven't found best way to assess coding skills

The data set `country_data` contains the life expectancy (in years) for different countries over several years. Below are the first few lines of the `country_data` data frame:

```
## # A tibble: 6 x 6
##   country    continent `1992` `1997` `2002` `2007`
##   <fct>      <fct>      <dbl> <dbl> <dbl> <dbl>
## 1 Afghanistan Asia         41.7  41.8  42.1  43.8
## 2 Albania     Europe        71.6  73.0  75.7  76.4
## 3 Algeria     Africa         67.7  69.2  71.0  72.3
## 4 Angola      Africa         40.6  41.0  41.0  42.7
## 5 Argentina   Americas       71.9  73.3  74.3  75.3
## 6 Australia   Oceania        77.6  78.8  80.4  81.2
```

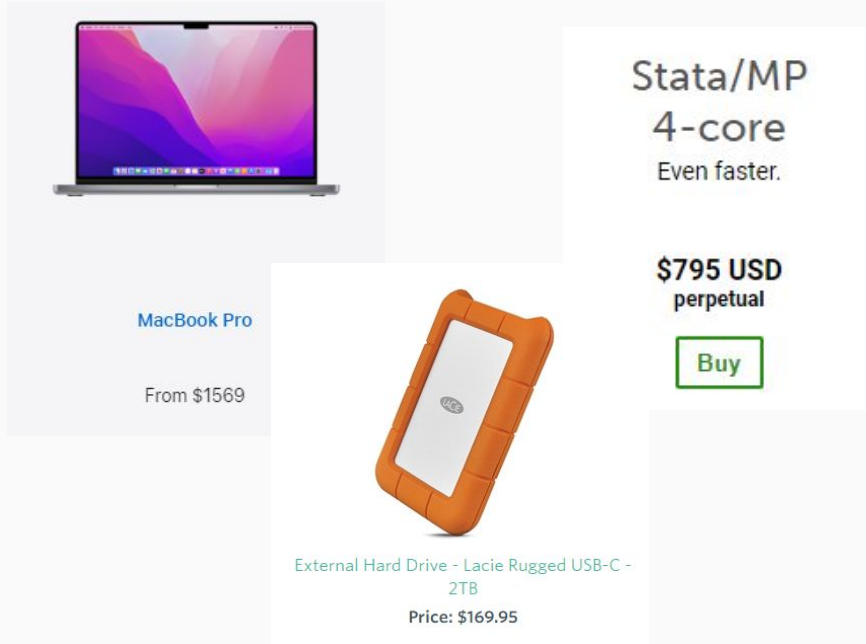
We first want to put the data in a tidy format, then we want to find the average life expectancy for each continent in the year 2007. Fill in the blanks (`...`) in the code below to accomplish that.

```
country_data %>%
  ... (cols = `1992`:`2007`,
        ... = "year",
        ... = "lifeExp") %>%
  ... (...) %>%
  ... (...) %>%
  ... (average = ...)
```

Example quiz question from DSCI100 using Canvas and R Exams

Research and Hardware Requirements

UBC Strategy 2.8: “UBC will expand opportunities for undergraduates to gain first-hand experience in research”



The image shows a screenshot of an Amazon product page. On the left, a MacBook Pro is displayed with a purple and pink abstract wallpaper. Below it, the text reads "MacBook Pro" and "From \$1569". To the right, a Lacie Rugged USB-C external hard drive is shown, which is orange and ruggedized. Below it, the text reads "External Hard Drive - Lacie Rugged USB-C - 2TB" and "Price: \$169.95". In the center-right, there is a product listing for "Stata/MP 4-core" with the text "Even faster.", "\$795 USD perpetual", and a green "Buy" button.

- The rapid expansion of data science tools has made these tools more accessible than ever before
- However, it also means that institutional supports have fallen away - replaced with personal items
- For projects - serious or pedagogical - hardware and software costs become significant barriers

This has the consequence that benefits tend to accrue to those most able to access them - which can exacerbate inequalities in access to data science education

Opportunities

Leveraging the Jupyter ecosystem to go all in: JupyterBook



DATA 301

Search this book...

Unsyllabus

ABOUT THIS COURSE

Course Syllabus (Official)

Course Schedule

GETTING STARTED!

Data Science Tech Stack

WEEKLY CONTENT

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6 (Break!)

Week 7

Week 8

Week 9

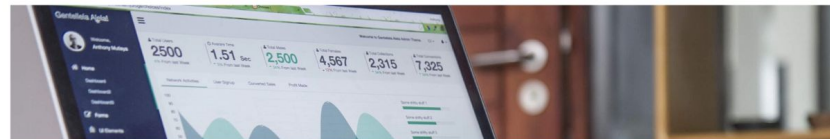
Week 10

Week 11

Week 12



Contents



THE UNIVERSITY OF BRITISH COLUMBIA

Irving K. Barber Faculty of Science
Okanagan Campus

Unsyllabus

Important Details

Name	Description
Course	DATA 301 or COSC 301
Term	2021 Winter Term 1
Instructor	Dr. Firas Moosvi (he/his/him)
Lectures	Monday, Wednesday, Friday from 2-3 PM: Online (Zoom) - Lectures will be recorded and available on Canvas.
Student Hours	To get live 1 on 1 help in the course, use Zoom at various times (see below for schedule).
Canvas URL	https://canvas.ubc.ca/courses/81264
Course Discussion	To ask any course-related questions, use private (personal, not useful for anyone else) or public (helpful for other) messages on Ed Discussion

Important Details

Contact Us

Evaluation

Passing requirements

Learning Intentions

Schedule

Getting Help

Syllabus vs. Unsyllabus

Unsyllabus changes

How do I do well in this course?

Course Tools

Teaching Philosophy

Academic Integrity

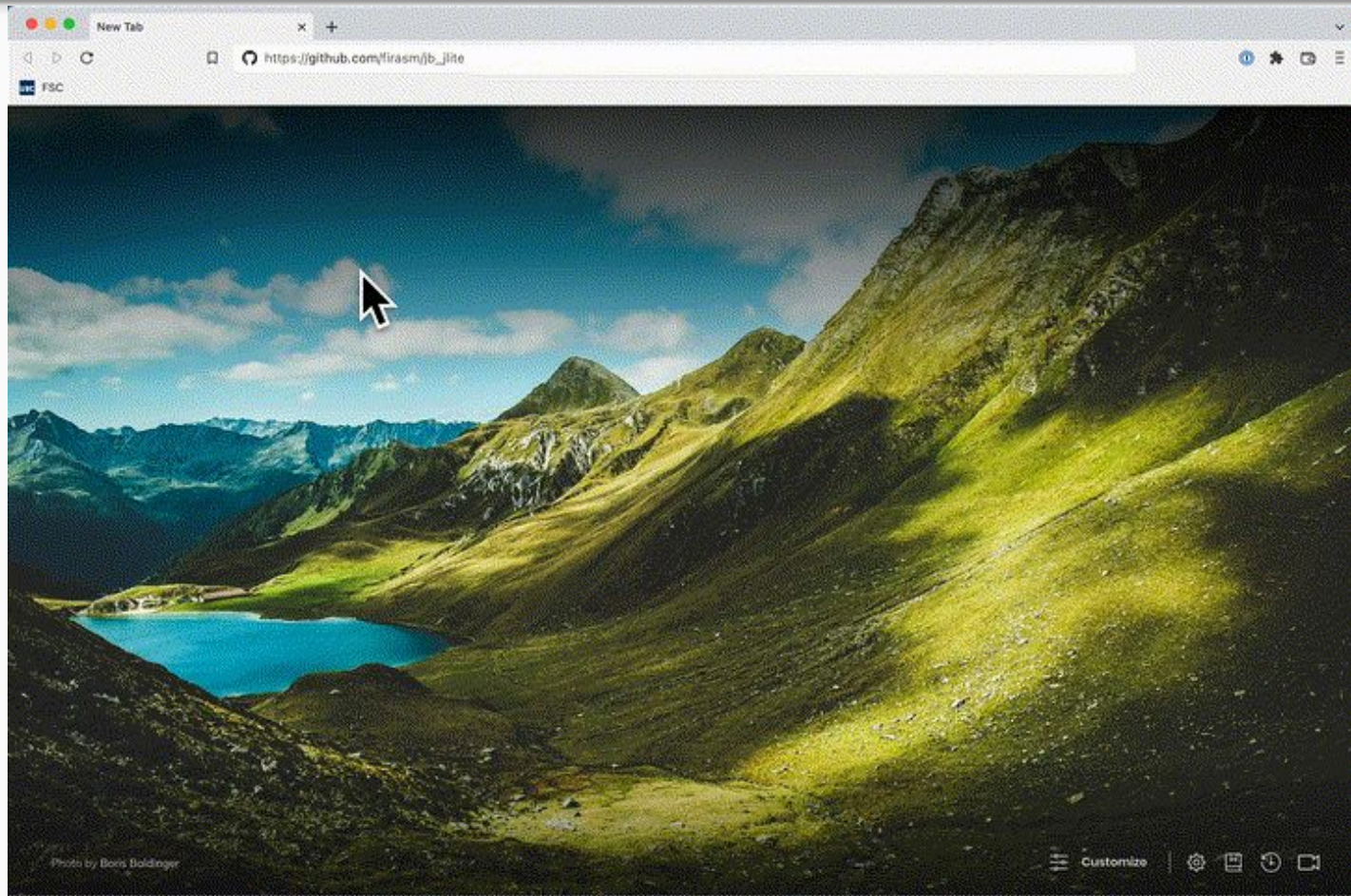
Course Accommodations

What should I think about if I'm considering withdrawing from the course?

Acknowledgements

References

JupyterLite: Programming in the cloud



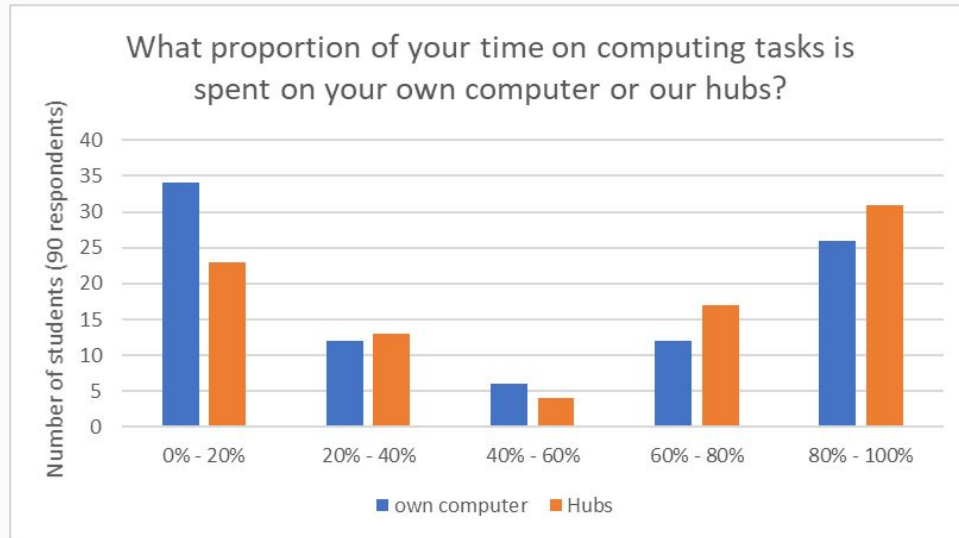
Student and Instructor Learning Technology Support

- Development of two Jupyter services: JupyterHub Open and JupyterHub Course
- JupyterHub Open: Compute resources available to all with browser access
 - Centrally supported and free for Students, Faculty, and Staff
- JupyterHub Course: Customizable, Canvas Integrated compute environment
 - Shared cost model with Departments and Faculties
- Allows for focussed central efforts
- Much lower bar for access



EOSC 211: from closed to open

- Transition funded as part of the OCESE TLEF
 - EOSC 211 jupyterbook and jupyterhub
 - EOSC 211 open source textbook
- All code and materials on github



DSCI100: testthat package in R

- Testthat package in R + Jupyter Notebooks
 - Instant, low-stakes feedback



testthat

Question 4.1

{points: 1}

Use the `min` function to find the minimum value of the numbers in the cell.

Assign the value to an object called `smallest`.

▶ `# Replace the fail() with your answer.`

`# your code here`

```
smallest <- min(2, 15, 4, 7)
smallest
```

2

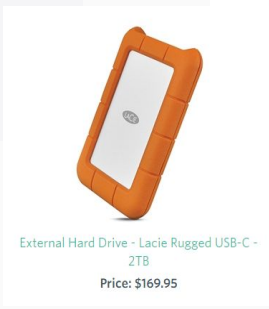
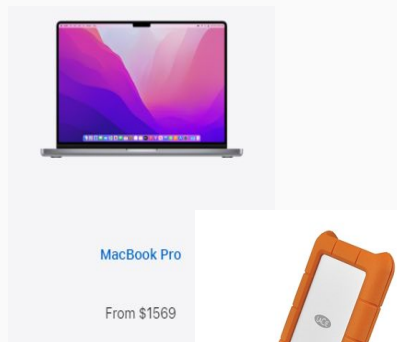
▶ `test_4.1()`

```
[1] "Success!"
```

```
test_4.1 <- function(){
  test_that('Solution is incorrect', {
    expect_equal(digest(smallest), 'db8e490a925a60e62212cefc7674ca02')
  })
  print("Success!")
}
```

Open Tools and Teaching

Replace proprietary tools with open-source, cloud-based frameworks and tools

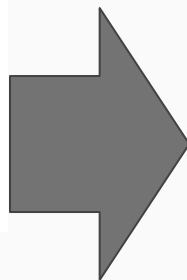


~\$2500

Stata/MP
4-core
Even faster.

\$795 USD
perpetual

Buy



~\$100

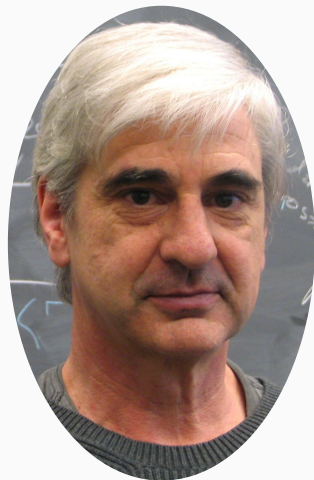


- Modern tools can now compete with closed source solutions
- Cloud-based tools like Jupyter eliminate costly hardware requirements
- Unifies research with teaching - experiential

Questions?

References

Contact Information



Phil Austin

paustin@eoas.ubc.ca



Jonathan Graves

jonathan.graves@ubc.ca



Melissa Lee

melissa.lee@stat.ubc.ca



Stephen Michaud

stephen.michaud@ubc.ca



Firas Moosvi

firas.moosvi@ubc.ca

UBC Data Science Slack:

https://join.slack.com/t/ubc-datascience/shared_invite/zt-194coq0ip-sch4vp0me0abgFwDEhpCbg

Resources

- [Jupyter Project](#)
- [Jupyter Book](#)
- [JupyterLite](#)
- [RMarkdown and Quarto](#)
- [DSCI 100 Homepage](#)
- [EOAS Project](#)